I'm not robot

When to use poisson pdf and cdf

When to use pdf and cdf statistics. Whats the difference between pdf and cdf. When do we use poisson distribution. When to use poisson distribution. What is normal pdf and cdf.

Random Variables and Probability Distributions — An Overview This journey isn't just for experienced statisticians but for anyone curious about learning. Introduction to key terms is essential before diving into the main content. Common Terms Used in the Blog ---------- Understanding these terms will help you follow along smoothly. Before we begin, let's cover some common terms used frequently in this blog. What are Random Variables? ---------- A random variable is a mathematical representation of the numerical outcomes from a random process or experiment. It represents variability found in random events like rolling dice or measuring traits of a group. Example: Rolling a fair six-sided die ---------- In this scenario, X represents the outcome of rolling the die. The possible values are 1 to 6 (inclusive), and each has an equal probability of occurring. If we define X as a random variable representing the number rolled on the die, its values represent possible outcomes of the random process. Types of Random Variables ---------- Random variables can be discrete or continuous, depending on whether they have a countable number of distinct values (e.g., coin flips) or an infinite range of values within a certain interval (e.g., height of individuals). Understanding random variables is essential in probability theory and statistics for analyzing and modeling uncertainty. Probability Distributions ---------- Probability distributions describe the probabilities associated with possible outcomes of a random variable. They specify how likely each outcome is to occur. A probability distribution assigns probabilities to each possible value that a random variable can take. Poisson distribution is a discrete probability distribution which describes the number of events occurring in a fixed interval of time or space. It is characterized by one parameter, λ, which represents the constant rate of events. The probability mass function (PMF) of a Poisson distribution is given by P(X=k), where k is the number of successes. Commonly used to model situations like coin flips, binomial distributions have applications in statistics, biology, economics and more. The distribution is defined by two parameters: n, the number of trials, and p, the probability of success in each trial. Let's explore the probability of customer arrivals in the next hour. We're assuming that the number of customers arriving follows a Poisson distribution with an average rate of 4 customers per hour. Here's how we can apply this distribution: The Probability Mass Function (PMF) gives us the likelihood of observing a specific number of customers within an hour. For instance, suppose we want to find the probability of exactly 3 customers arriving in the next hour. We'd use the PMF with λ = 4 and k = 3, which yields approximately 0.1952 or 19.52%. Now, let's switch gears to the normal distribution, also known as the Gaussian distribution. This continuous distribution is symmetric around its mean, resembling a bell curve. The Probability Density Function (PDF) provides the probability density at each point but doesn't directly give us the probability of a specific outcome. Instead, we integrate the PDF over a certain range to find the probability. Some key characteristics of the normal distribution include: * The mean, median, and mode are always equal. * The total area under the curve is always 1. * The distribution is symmetric around its mean. The relationship between probability and the PDF is straightforward. While the PDF provides the density of probability at each point, it doesn't directly provide the probability of a specific outcome occurring. Instead, we integrate the PDF over that range to find the desired probability. Lastly, the Empirical Rule states that in many cases, data follows a normal distribution. According to this rule: * Approximately 68% of the data falls within one standard deviation of the mean. * Approximately 95% of the data falls within two standard deviations of the mean. * Approximately 99.7% of the data falls within three standard deviations of the mean. For a standard normal distribution (mean μ = 0 and standard deviation σ = 1), the PDF is given by the Gaussian function, which we'll explore further in future calculations. Consider IQ scores assumed to follow a normal distribution with mean 100 and standard deviation 15. With a population of 1000 individuals, applying the empirical rule reveals: - Approximately 680 will have IQs between 85 and 115. - About 950 will have IQs between 70 and 130. - Roughly 997 will have IQs between 55 and 145. A standard normal distribution (z) allows us to normalize any normal distribution by subtracting the mean and dividing by the standard deviation. This process, called normalization, enables comparison of data across different scales. Imagine two variables: one for income in dollars and another for age in years. By normalizing the data, we can transform both onto a standardized scale (z-scores), making them more comparable for analysis. The uniform distribution is a probability distribution where all outcomes are equally likely within a given range. It's used in various applications like modelling random processes, generating random numbers, and certain sampling techniques. Think of drawing a slip from a box containing numbered slips from 1 to 10 – each has an equal chance of being drawn. The Cumulative Distribution Function (CDF) provides the probability that a random variable is less than or equal to a specified point. When using the CDF, note that it calculates the probability of a range less than or equal to the specified value by default. it by subtracting the CDF value from 1. Now, let's talk about the Percentage Point Function (PPF). This fancy term refers to the math behind finding out what value a certain percentage of observations fall below. Think of it like this: if you have a bunch of data and want to know where the "cut-off" point is for a specific percentage, say 95%, the PPF would give you that magic number. In simpler terms, the PPF is like a "what's the cutoff?" function. For instance, imagine you have a bunch of scores on a test and you want to know what score marks the top 50%. The PPF would tell you exactly where that score lies. The PPF is actually the opposite of something called the Cumulative Distribution Function (CDF). While the CDF shows you the probability that a random variable is less than or equal to a certain point, the PPF flips it around and tells you what point corresponds to a specific probability. It's like having a "probability-to-point" translator. Now, there are some super useful tools in Python called scipy.stats that can help us work with all these concepts. With these tools, we can easily calculate probabilities for various types of distributions, such as the normal distribution, binomial distribution, uniform distribution, and many more. To summarize, when working with random variables, we have PDFs (probability density functions) and CDFs (cumulative distribution functions), which are like "point-to-probability" translators. The PPF is similar but works in reverse, telling us what point corresponds to a specific probability. It's all about understanding these concepts so we can work with data and statistics with confidence! **Continuous Random Variables** A continuous random variable can take on an infinite number of values, such as a person's height, weight, or time to run a mile. For instance, a person's height could be 60.2 inches, 65.2344 inches, or any other value within a given range. **Discrete vs. Continuous Variables** If you can count the possible outcomes of a random variable, it is considered discrete (e.g., counting the number of times a coin lands on heads). However, if you can measure the outcome, such as height or weight, it is considered continuous. **Probability Density Functions (PDFs)** A PDF describes the probability that a random variable takes on a specific value. For example, rolling a dice once and letting x denote the number landed on, we have: P(x = 0) = 1/6 P(x = 1) = 1/6 ... P(x > 6) = 0 **Continuous PDFs** However, for continuous random variables, such as weight or height, it is impossible to assign a non-zero probability to any exact value. For instance, the probability that a burger weighs exactly 0.25 pounds is essentially zero. **Cumulative Distribution Functions (CDFs)** A CDF describes the probability that a random variable takes on a value less than or equal to x. Using the dice example again: P(x ≤ 0) = 0 P(x ≤ 1) = 1/6 ... P(x > 6) = 0 Note that for continuous variables, CDFs are often used in place of PDFs due to the infinite number of possible values. A cumulative distribution function (cdf) has two key properties: it assigns zero probability to the smallest possible value of a random variable, and one probability to the largest possible value. The probabilities in between are non-decreasing, meaning that as the value increases, the probability also increases. This is demonstrated by the example of a dice rolling, where the probability of landing on each number from 1 to 6 is cumulative and non-decreasing. The relationship between a cdf and a probability density function (pdf) is closely tied, with the pdf being the derivative of the cdf. This means that the area under the curve of the pdf between negative infinity and x equals the value of x on the cdf.